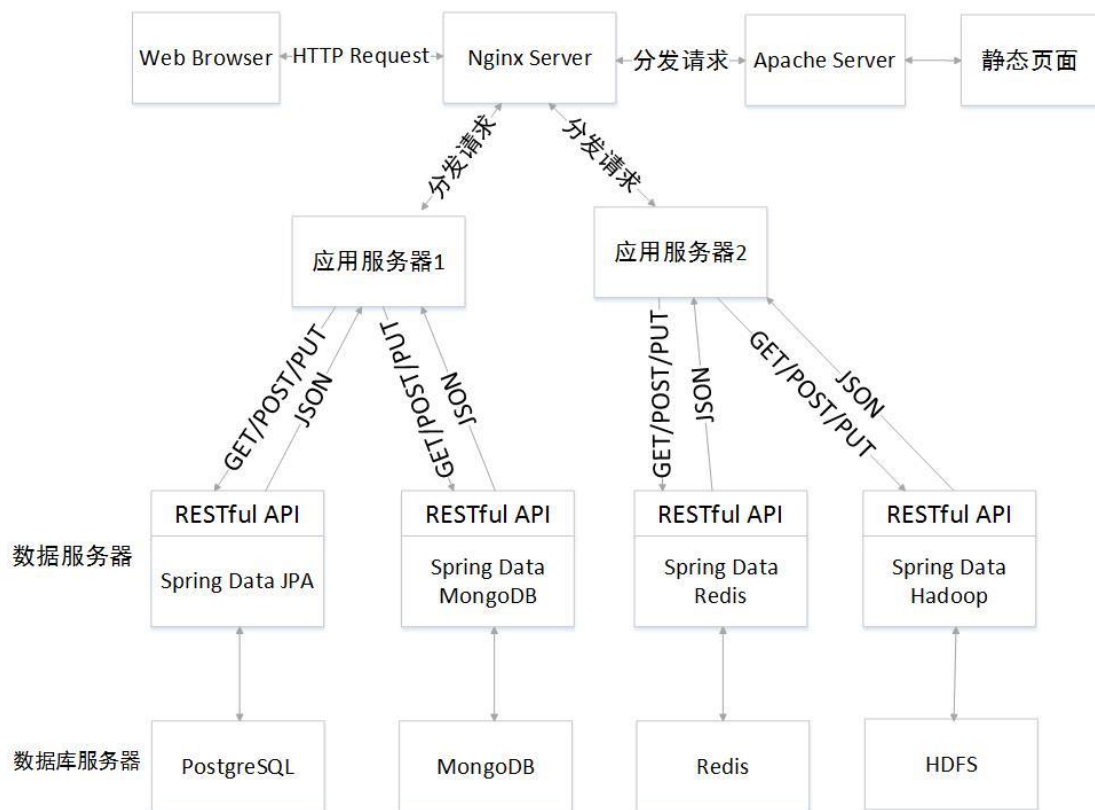


本周周报（6.8-6.14）

刘昊南

本周工作

1. 本周调研了大数据混合持久化层的编程实现上的解决方案
2. 因为 Java 的跨平台特性以及丰富的后端框架，以及各个数据库均提供 Java Driver，与 Hadoop 和 Spark 交互的便利，初步决定使用 Java 来搭建数据库的服务层
3. 数据服务调用接口：采用 RESTful 服务的方式提供数据访问层的调用接口，在三种主流的 Web 服务实现方案中，因为 REST 模式与复杂的 SOAP 和 XML-RPC 相比更加简洁，越来越多的 web 服务开始采用 REST 风格设计和实现。例如，Amazon.com 提供接近 REST 风格的 Web 服务进行图书查找；雅虎提供的 Web 服务也是 REST 风格的。
4. 数据传输：采用 JSON 表达数据服务返回的内容，JSON 是一种轻量级的数据交换语言，以文字为基础，且易于让人阅读，在 RESTful 的 web 服务中使用较多，相比于 XML 传输效率更高
5. 数据访问层框架：采用 spring data family 持久层框架来搭建，spring data 提供了与关系数据库、非关系数据库、map-reduce 框架的数据访问抽象
 - a) 关系数据库：采用 spring data jpa 实现关系型数据库的数据访问层
 - b) 非关系数据库：采用 spring data mongodb 实现 MongoDB 的数据访问层，采用 spring data redis 实现 Redis 的数据访问
 - c) HDFS：采用 spring data Hadoop 实现对 HDFS 中文件的数据访问
6. 目前构想的具体实现方案如下图所示：
 - a) 前端采用 Nginx 服务器来分发请求并做负载均衡，将从浏览器端发来的 http 请求分发给 apache 服务器和应用服务器
 - b) apache 服务器负责托管静态的 html 页面
 - c) 应用服务器托管动态页面如 php、jsp 等，并处理数据访问和数据计算的请求，即后端应用层的逻辑代码在应用服务器执行，不同的应用服务（如 vaud 和农业大数据）托管在不同的应用服务器，由 nginx 负责将对应的请求转发给对应的服务器执行，应用服务器可以有冗余，由 nginx 做好负载均衡即可
 - d) 数据服务器对外提供 RESTful API 供应用服务器进行通信，应用服务器可以通过 GET 和 POST 请求来获取数据，通过 POST 请求来修改数据，通过 PUT 请求来插入新数据，数据服务器以 JSON 的形式来返回数据
 - e) 数据服务器通过 Spring Data 数据访问层框架来访问 PostgreSQL 等关系型数据和 MongoDB、Redis 等非关系型数据库的数据和 HDFS 分布式文件系统中的文件



下周计划

1. 为 vaud 搭建初步的数据访问服务